Part A:

Logistic Regression Model
Model Fitting and Prediction:

We fitted a logistic regression model with the training set. On the test set, predictions were made, and a threshold of 0.5 was used to convert the results to binary outcomes.

Model Evaluation:

```
## Logistic Regression Metrics://n
```

```
cat("Accuracy:", accuracy_logistic, "//n")
```

```
## Accuracy: 0.910327 //n
```

```
cat("Precision:", precision_logistic, "//n")
```

```
## Precision: 0.9296425 //n
```

```
cat("Recall:", recall_logistic, "//n")
```

```
## Recall: 0.9725465 //n
```

```
cat("F1 Score:", f1_score_logistic, "//n")
```

```
## F1 Score: 0.9506107 //n
```

```
cat("AUC:", auc_logistic, "//n//n")
```

```
## AUC: 0.9369781 //n//n
```

The logistic regression model performs very well across all metrics, with high accuracy, precision, recall, F1 score, and AUC. This suggests the model is effective at making accurate predictions, capturing true positive cases, and distinguishing between classes.

Logistic Regression Coefficients:

- The summary of the logistic regression model provided the coefficients, standard errors, z-values, and p-values.
- Key findings included significant positive coefficients for variables like "jobretired," "educationuniversity.degree," "poutcomesuccess," and "duration," indicating these factors positively influence the likelihood of a successful outcome.
- Negative coefficients were observed for "jobblue-collar," "defaultunknown," "contacttelephone," and "emp.var.rate," suggesting these variables decrease the likelihood of a successful outcome.

Decision Tree Model

Model Fitting and Prediction:
A decision tree model was fitted using the training data. Predictions were made on the test set.
Model Evaluation:

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    no   yes
##        no  10581   714
##        yes   383   678
##
##                Accuracy : 0.9112
##                  95% CI : (0.9061, 0.9162)
##     No Information Rate : 0.8873
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.5045
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.9651
##             Specificity : 0.4871
##          Pos Pred Value : 0.9368
##          Neg Pred Value : 0.6390
##              Prevalence : 0.8873
##          Detection Rate : 0.8563
##    Detection Prevalence : 0.9141
##       Balanced Accuracy : 0.7261
##
##        'Positive' Class : no
##
```

Interpretation:

- The model shows high accuracy (91.12%) and high sensitivity (96.51%), indicating it captures most of the true positives.
- Precision is also high (93.68%), suggesting the model makes few false positive errors.
- Specificity (48.71%) and Negative Predictive Value (63.90%) are relatively lower, indicating the model has more difficulty correctly identifying true negatives.
- Overall, the model performs well in predicting the positive class but could be improved in distinguishing true negatives.

Comparison of Models

- Both models achieved similar accuracy, with the decision tree model slightly outperforming the logistic regression model.
- The logistic regression model had a higher AUC , indicating better overall performance in distinguishing between classes.
- Both models showed high sensitivity, with the logistic regression model having a slightly higher sensitivity .
- The decision tree model had a higher specificity.

Visualizations:

AUC-ROC curves were plotted for both models, visually comparing their performance.
The structure of the decision tree was also visualized to understand the decision rules derived by the model.

Conclusion

- The logistic regression model and decision tree model both performed well, with the logistic regression model demonstrating slightly better discrimination ability as indicated by a higher AUC.
- The decision tree model provided better interpretability and slightly higher accuracy.
- Depending on the specific application needs (e.g., ease of interpretation vs. model performance), either model could be considered appropriate for predicting the target variable in this dataset.